INFORMATION TECHNOLOGY AND SYSTEMS CENTER
UNIVERSITY OF ALABAMA IN HUNTSVILLE
HUNTSVILLE, AL-35899

AN INTEROPERABLE FRAMEWORK FOR MINING AND ANALYSIS OF SPACE SCIENCE DATA (F-MASS)

LEAD
DR. SARA GRAVES
DR. RAHUL RAMACHANDRAN
DR. XIANG LI

ANNUAL REPORT

May 1, 2004

PROJECT OBJECTIVES

The objective of this project is to extend the existing ADaM system to create a Framework for Mining and Analysis of Space Science (F-MASS) data. Reaching this objective will involve an investigation of data sets, data processing methodologies and analysis procedures unique to the space science community. Specific space science research problems will be targeted to guide extensions to the ADaM system to support the community's needs. Such new development will include enhancing selected ADaM modules for application to space science problems, and integrating new modules into the existing framework. These new modules will include input and output filters for space science data, mining modules to address the target research problems, and a customized user interface. The resulting software suite, along with user documentation will be made openly available to space science researchers.

YEAR TWO MILESTONES AND OBJECTIVES

- **Identify and implement space science/data mining algorithms**

The task for the project team identified in the proposal was to search, survey and identify data mining algorithms suitable for space science research. In particular, the research team sought for the algorithms for the use case studies proposed in this research project.

- **Develop F-MASS data input modules**

The task for the project team was to develop data input modules for selected data sets for a given science scenario. The *input module has been designed and implemented in year one* and reported in year one annual report.

- **Design and implement space science user interface**

This user interface will allow space scientists to fully exploit the various data mining algorithms provide by the F-MASS system.

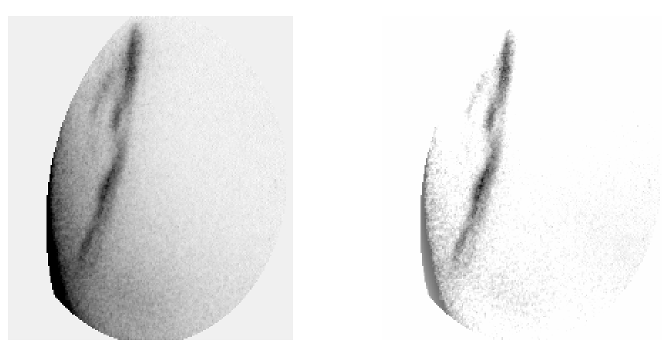
TASKS ACCOMPLISHED IN YEAR II

- **Identification and implementation of space science/data mining algorithms based on selected scenarios: Polar cap boundary and auroral oval identification**

The polar cap boundary and aurora identifications are the first two of the three case study scenarios proposed in the research project. The auroral oval is one of the significant phenomena in the polar region and the study of aurora and polar cap is a key research area in understanding the solar-terrestrial connections and interactions, as indicated in the research proposal. The polar boundary is the pole-ward boundary of an auroral oval, constantly locating above the polar zone. The auroral oval and polar cap are the areas for various types of aurora and sub-storm study. Initial effort has been put in year one for the polar cap boundary detection using the UVI measurement on board the POLAR satellite. Three methods have been examined in year one for polar cap boundary detection: 1) local thresholding, 2) Heuristic Method, and 3) K-Means Clustering.

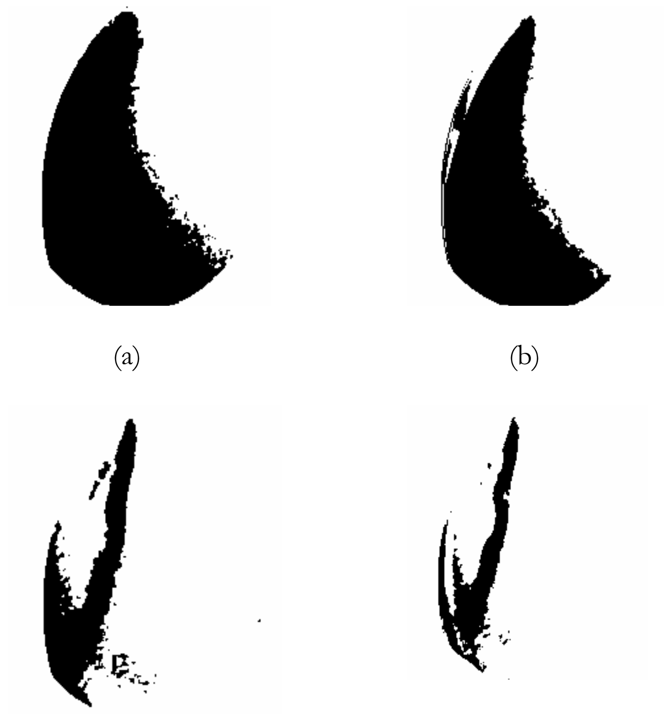
Although several kinds of methods have been examined in the past for auroral oval detection, including neural network and elastic curve methods, thresholding methods are still the most intuitive and effective method. In year two of the project, the research team continued to examine several kinds of thresholding algorithms. The examined algorithms include the *mixture modeling* thresholding algorithm, and algorithms based on *the fuzzy set* and *entropy*. Also two thresholding approaches were examined: global and adaptive (local) approach. The performances of the three kinds of thresholding algorithms with two approaches were evaluated and compared using visual inspection. As expected, adaptive approach performed better than global approach. However, adaptive approach may present a problem of discontinuity of auroral oval boundaries. **These results were presented in the IEEE ITCC 2004 conference. The details can be found in the attached paper**.

As it was found in the analysis, the performance of the auroral oval detection is significantly deteriorated with the existence of dayglow emission in the satellite images. Even adaptive thresholding technique fails to detect auroral ovals with strong dayglow emission. The dayglow emission may exist in the UVI images all the time except the winter season. The dayglow effect is the strongest in the summer season. To make full use of UVI images in aurora study, dayglow effect has to be removed from the image intensity in order to detect the auroral ovals accurately. Therefore, the second research task in year two was for dayglow removal process. A dayglow model was developed as function of solar and satellite viewing angles using regression technique. The dayglow model can be used to estimate the dayglow intensity and removed from the image. **The results for dayglow removal research were submitted to the 2004 IGARSS conference**. Figure 1 shows the performance of dayglow removal for UVI LBHL band image at 04:28:36 on July 20, 2000. Before dayglow removal, the dayglow emission can be seen on the lower left part of the image (lower left part of background image is darker than that of upper right part). Figure 2 shows the improvement of auroral oval detection after dayglow removal using mixture modeling algorithm. From Figure 2, the algorithm failed to detect aurora oval for both global and adaptive thresholding techniques even though visually the aurora was clearly seen. After dayglow removal, the aurora oval was detected using both global and adaptive thresholding techniques.



(a) (b)

Figure 1. (a) UVI LBHL band image at 04:28:36 on July 20, 2000
(b) UVI image of (a) after dayglow removal



(a) (b)

(c) (d)

Figure 2. Results of auroral oval detection using mixture modeling algorithm for UVI image at 04:28:36 on July 20, 2000. (a) global thresholding without dayglow removal, (b) adaptive thresholding without dayglow removal, (c) global thresholding with dayglow removal, (d) adaptive thresholding with dayglow removal

The current effort of the project is focused on refining the auroral oval detection using these thresholding techniques. Furthermore, another thresholding (Otsu) algorithm was also included in our design, evaluation and implementation. From the UVI images examined, it seems that in general the mixture modeling algorithm underestimates the optimal threshold value while the Otsu algorithm over-estimates the optimal threshold value. The ensemble of the two thresholding algorithms may perform better in auroral oval detection and will be explored in Year III. The research team examined the post-processing methods to filter out the noise in the detected aurora oval as shown in Figure 2 (c) and 2(d) and expect to develop a fully automated procedure that can detect aurora ovals reasonably well under most of circumstances all year round.

- **User Interface Design and Development of F-MASS**

For F-MASS, the original ADaM was redesigned as a mining toolkit, with each algorithm available as a stand alone executable. In addition, Python interfaces were added to these stand alone executables. Python is a very simple and easy to learn programming language and is used by a large portion of the science community. These Python module user interfaces to the F-MASS toolkit allows scientists to call any of the mining operations directly into their programs, script together complex mining operations and fully exploit full capabilities of a programming language (loops, if conditions) within mining scripts.

PUBLICATIONS IN YEAR II

- Xiang Li, Rahul Ramachandran, Sunil Movva, Sara Graves, Glynn Germany, Wladislaw Lyatsky, A. Tan, 'Dayglow removal from FUV auroral images', Submitted to International Geoscience and Remote Sensing Symposium (IGARSS), 2004

- X. Li, R. Ramachandran, M. He, S. Movva, J. Rushing and S. Graves, Comparing Different Thresholding Algorithms for Segmenting Auroras, Space Science Computation and IT Applications , International Conference on Information Technology, Las Vegas, NV, 2004.